

Cryptanalysis of an Encryption Scheme Based on Blind Source Separation

Shujun Li, Chengqing Li, Kwok-Tung Lo, *Member, IEEE* and Guanrong Chen, *Fellow, IEEE*

Abstract—Recently Lin et al. proposed a method of using the underdetermined BSS (blind source separation) problem to realize image and speech encryption. In this paper, we give a cryptanalysis of this BSS-based encryption and point out that it is not secure against known/chosen-plaintext attack and chosen-ciphertext attack. In addition, there exist some other security defects: low sensitivity to part of the key and the plaintext, a ciphertext-only differential attack, divide-and-conquer (DAC) attack on part of the key. We also discuss the role of BSS in Lin et al.’s efforts towards cryptographically secure ciphers.

Index Terms—blind source separation (BSS), speech encryption, image encryption, cryptanalysis, known-plaintext attack, chosen-plaintext attack, chosen-ciphertext attack, differential attack, divide-and-conquer (DAC) attack.

I. INTRODUCTION

With the rapid development of multimedia and networking technologies, the security of multimedia data becomes more and more important in many real applications. To fulfill such an increasing demand, during past decades many encryption schemes have been proposed to protect multimedia data, including speech, images and videos [1]–[9].

According to the nature of protected data, multimedia encryption schemes can be classified into two basic types: analog and digital. Most early schemes were designed to encrypt analog data in various ways: element permuting, signal masking, frequency shuffling, etc., all of which may be exerted in time domain or transform domain or both. However, due to the simplicity of the encryption procedures, almost all analog encryption schemes are not sufficiently secure against cryptographical attacks, especially those modern attacks such as known/chosen-plaintext and chosen-ciphertext attacks [2], [3], [10], [11]. As a comparison, in digital encryption schemes, one can employ any cryptographically strong cipher, such as DES [12] or AES [13], to achieve a higher level of security. Besides, to achieve a higher efficiency of encryption and some special demands of multimedia encryption (such as format-compliance [14] and perceptual encryption [15]), many specific multimedia encryption schemes have also been

This research was partially supported by The Hong Kong Polytechnic University’s Postdoctoral Fellowships Program under grant no. G-YX63. The work of K.-T. Lo was supported by the Research Grants Council of the Hong Kong SAR Government under Project Number 523206 (PolyU 5232/06E).

Shujun Li and Kwok-Tung Lo are with the Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong SAR, P. R. China.

Chengqing Li and Guanrong Chen are with the Department of Electronic Engineering, City University of Hong Kong, Kowloon Toon, Hong Kong SAR, P. R. China.

The corresponding author is Shujun Li. Contact him via his person web site: <http://www.hooklee.com>.

developed [4]–[6]. Recent cryptanalysis work [16]–[30] has shown that some multimedia encryption schemes are insecure against various cryptographical attacks.

Recently Lin et al. suggested employing blind source separation (BSS) for the purpose of image and speech encryption [31]–[37]. The basic idea is to mix multiple plaintexts (or multiple segments of the same plaintext) with a number of secret key signals, in the hope that an attacker has to solve a hard mathematical problem – the underdetermined BSS problem. In Sec. VII of [37], Lin et al. claimed that this BSS-based cipher “is immune from the attacks such as the ciphertext-only attack, the known-plaintext, and the chosen-plaintext attack”, “as long as the intractability of the underdetermined BSS problem is guaranteed by the mixing matrix for encryption”.

This paper re-evaluates the security of the BSS-based encryption scheme and points out that it is actually insecure against known/chosen-plaintext attack and chosen-ciphertext attack. In addition, some other security defects are also found under the ciphertext-only attacking scenario, including the low sensitivity to the mixing matrix (part of the secret key) and the plaintext, and a differential attack that works well when the matrix size is small. Based on the cryptanalytic findings, we also discuss the role of BSS in Lin et al.’s efforts towards cryptographically secure ciphers.

The rest of this paper is organized as follows. In next section we give a brief introduction to the BSS-based encryption scheme. Section III is the main body of this paper and focuses on the cryptanalysis of the BSS-based encryption scheme. Then, the role of BSS in cryptography is discussed in Sec. IV. Finally the last section concludes this paper.

II. BSS-BASED ENCRYPTION

Blind source separation is a technique that tries to recover a set of unobserved sources or signals from observed mixtures [38]. Given N unobserved signals s_1, \dots, s_N and a mixing matrix \mathbf{A} of size $N \times M$, the BSS problem is to recover s_1, \dots, s_N from M observed signals x_1, \dots, x_M , where

$$[x_1, \dots, x_M]^T = \mathbf{A}[s_1, \dots, s_N]^T. \quad (1)$$

When $M \geq N$, the blind source separation is possible when \mathbf{A} satisfies some requirements. However, when $M < N$, this is generally impossible (whatever \mathbf{A} is), thus leading to the underdetermined BSS problem.

In [31]–[37], Lin et al. introduced a number of secret key signals to make the determination of the plaintext signals become an underdetermined BSS problem in the case that the key signals are unknown. Given P input plain-signals

$s_1(t), \dots, s_P(t)$ and Q key signals $k_1(t), \dots, k_Q(t)$, the encryption procedure is described as follows¹:

$$\mathbf{x}(t) = [x_1(t), \dots, x_P(t)]^T = \mathbf{A}\mathbf{s}_k(t), \quad (2)$$

where $\mathbf{x}(t)$ denote P cipher-signals, $\mathbf{s}_k(t) = [s_1(t), \dots, s_P(t), k_1(t), \dots, k_Q(t)]^T$, and \mathbf{A} is a $P \times (P+Q)$ mixing matrix whose elements are within in $[-1, 1]$. Assume that $\mathbf{A} = [\mathbf{A}_s, \mathbf{A}_k]$, where \mathbf{A}_s is a $P \times P$ matrix and \mathbf{A}_k is a $P \times Q$ matrix. Then, the encryption procedure can be represented in an equivalent form:

$$\mathbf{x}(t) = \mathbf{A}_s\mathbf{s}(t) + \mathbf{A}_k\mathbf{k}(t), \quad (3)$$

where $\mathbf{s}(t) = [s_1(t), \dots, s_P(t)]^T$ and $\mathbf{k}(t) = [k_1(t), \dots, k_Q(t)]^T$. Thus, as long as \mathbf{A}_s is an invertible matrix, one can decrypt $\mathbf{s}(t)$ as follows²:

$$\mathbf{s}(t) = \mathbf{A}_s^{-1}(\mathbf{x}(t) - \mathbf{A}_k\mathbf{k}(t)). \quad (4)$$

Different values of Q was used in Lin et al.'s papers: $Q = 1$ in [31] and $Q = P$ in [32]–[37]. When $Q = P$, Lin et al. further set $\mathbf{A}_s = \mathbf{B}$ and $\mathbf{A}_k = \beta\mathbf{B}$, where $\beta \geq 10$ for image encryption and $\beta \geq 1$ for speech encryption. In this case, the encryption procedure becomes

$$\mathbf{x}(t) = \mathbf{B}(\mathbf{s}(t) + \beta\mathbf{k}(t)), \quad (5)$$

and the decryption procedure becomes

$$\mathbf{s}(t) = \mathbf{B}^{-1}\mathbf{x}(t) - \beta\mathbf{k}(t). \quad (6)$$

Observing Eq. (3), one can see that the encryption procedure contains two steps:

- Step 1: $\mathbf{x}^{(1)}(t) = \mathbf{A}_s\mathbf{s}(t)$;
- Step 2: $\mathbf{x}(t) = \mathbf{x}^{(1)}(t) + \mathbf{A}_k\mathbf{k}(t)$.

The first step corresponds to a substitution (block) cipher, and the second step corresponds to a additive stream cipher. From another point of view, the two steps are exchanged as follows:

- Step 1: $\mathbf{x}^{(1)}(t) = \mathbf{s}(t) + \mathbf{A}_s^{-1}\mathbf{A}_k\mathbf{k}(t)$;
- Step 2: $\mathbf{x}(t) = \mathbf{A}_s\mathbf{x}^{(1)}(t)$.

In any case, the BSS-based encryption scheme is always a product cipher composed by a simple block cipher and an additive stream cipher. In next section, we will show that the two sub-ciphers can be separately broken by known/chosen-plaintext attack and chosen-ciphertext attack.

In the BSS-based encryption scheme, the key signals $k_1(t), \dots, k_Q(t)$ are as long as the plain-signals and have to be generated by a pseudo-random number generator (PRNG) with a secret seed I_0 , which serves as the secret key. In Lin et al.'s papers, it was not explicitly mentioned whether or not the mixing matrix should be used as part of the secret key. However, if the attacker knows \mathbf{A} , the product cipher degrades to be a stream cipher. Considering $\mathbf{x}^*(t) = \mathbf{A}_s^{-1}\mathbf{x}(t)$ as the equivalent cipher-signal, the encryption procedure becomes

$$\mathbf{x}^*(t) = \mathbf{s}(t) + \mathbf{A}_s^{-1}\mathbf{A}_k\mathbf{k}(t). \quad (7)$$

¹To achieve a clearer description of the BSS-based encryption scheme, in this paper we use some notations different from those in Lin et al.'s original papers. For example, in [37], the i -th key signal is denoted by $s_{ni}(t)$, while in this paper we use $k_i(t)$ to emphasize the fact that it is a **key** signal.

²In Lin et al.'s papers, it is said that the decryption procedure was achieved via BSS. However, from the cryptographical point of view, it is more convenient to denote the decryption procedure by Eq. (4).

In this case, the encryption scheme is actually independent of the underdetermined BSS problem. In addition, as we shown later in Sec. III-A.5, the key signals can be totally circumvented in a ciphertext-only differential attack, so the mixing matrix \mathbf{A} must be kept as the secret key. Thus, in this paper we assume that the secret key consists of both I_0 and \mathbf{A} .

In [31]–[35], the BSS-based encryption scheme was mainly designed to encrypt P images simultaneously, where $s_i(t)$ is the t -th pixel in the i -th image. In [36], [37], the encryption scheme was suggested to encrypt a single speech, each frame of which is divided into P segments and $s_i(t)$ is the t -th sample in the i -th segment. This encryption scheme can also be applied for a single image, by dividing it into P blocks of the same size. To facilitate the following discussion, we assume that the encryption scheme is used to encrypt a single plaintext with P segments of equal size.

In Sec. VII of [37], Lin et al. claimed that the BSS-based encryption scheme is secure against most modern cryptographical attacks, including the ciphertext-only attack, the known-plaintext attack, and the chosen-plaintext attack. In next section we will show that this claim is problematic.

III. CRYPTANALYSIS

Before introducing the cryptanalytic results, let us see how large the key space is. In Lin et al.'s papers, each element of \mathbf{A} is within the interval $[-1, 1]$. Then, assuming that each element in \mathbf{A} has R possible values³, the number of all possible mixing matrix \mathbf{A} is $R^{P(P+Q)}$. Furthermore, assuming that the bit size of I_0 is L , the size of the whole key space is $R^{P(P+Q)}2^L$. When $Q = P$ and $\mathbf{A} = [\mathbf{B}, \beta\mathbf{B}]$, the size of the whole key space is $R^{P^2}2^L$. Later we will show that the real size of the key space is much smaller than this estimation, due to some essential security defects of the BSS-based encryption scheme. We will also point out that the encryption scheme under study is not secure against known/chosen-plaintext attack and chosen-ciphertext attack.

A. Ciphertext-Only Attack

1) *Divide-and-Conquer (DAC) Attack*: Rewriting Eq. (4) in the following form:

$$\mathbf{s}(t) = \hat{\mathbf{A}}\mathbf{x}_k(t), \quad (8)$$

where $\mathbf{x}_k(t) = [x_1(t), \dots, x_P(t), k_1(t), \dots, k_Q(t)]^T$ and

$$\hat{\mathbf{A}} = \mathbf{A}_s^{-1}[\mathbf{I}, -\mathbf{A}_k] = [\mathbf{A}_s^{-1}, -\mathbf{A}_s^{-1}\mathbf{A}_k].$$

From the above equation, to recover $x_i(t)$, one only needs to know $\mathbf{k}(t)$ and the i -th row of $\hat{\mathbf{A}}$. In other words, when the BSS-based encryption scheme is used to encrypt P independent plaintexts, the i -th plaintext can be exactly recovered with the knowledge of I_0 and the i -th row of $\hat{\mathbf{A}}$. A similar result

³The value of R is determined by the finite precision under which the cryptosystem is realized. For example, if the cryptosystem is implemented with n -bit fixed-point arithmetic, $R = 2^n$; if it is implemented with IEEE floating-point arithmetic, $R \approx 2^{31}$ (single-precision) or $R \approx 2^{63}$ (double-precision) [39], where note that the sign bit of the floating-point number is always negative.

can be obtained when P segments of one single plaintext is encrypted with the encryption scheme. This fact means that P rows of $\hat{\mathbf{A}}$ can be separately broken with a divide-and-conquer (DAC) attack. As a result, the size of the key space is reduced to be $PR^{(P+Q)}2^L$. When $Q = P$ and $\mathbf{A} = [\mathbf{B}, \beta\mathbf{B}]$, it becomes PR^P2^L .

2) *Low Sensitivity to \mathbf{A}* : From the cryptographical point of view, given two distinct keys, even if their difference is the minimal value under the current finite precision, the encryption and decryption results of a good cryptosystem should still be completely different. In other words, this cryptosystem should have a very high sensitivity to the secret key [12]. Unfortunately, the BSS-based encryption scheme does not satisfy this security principle, because the involved matrix computation is not sufficiently sensitive to matrix mismatch. Given two matrices \mathbf{A}_1 and \mathbf{A}_2 of size $M \times N$, if the maximal difference of all elements is ε , then one can easily deduce that each element of $|\mathbf{A}_1\mathbf{s}(t) - \mathbf{A}_2\mathbf{s}(t)|$ is not greater than $N \max(\mathbf{s}(t))\varepsilon$. As a result, the matrix \mathbf{A} can be approximately guessed under a relatively large finite precision ε , still maintaining an acceptable quality of the recovered plaintexts. This immediately leads to a significant reduction of the size of the key space: from $PR^{(P+Q)}2^L$ to $P[2/\varepsilon]^{(P+Q)}2^L$, where $[2/\varepsilon]^{(P+Q)} \ll R^{(P+Q)}$.

The above low sensitivity can be easily verified with experiments described as follows:

- *Step 1*: for a randomly-generated key $(\mathbf{A}, \mathbf{I}_0)$, calculate the ciphertext $\mathbf{x}(t)$ corresponding to a plaintext $\mathbf{s}(t)$;
- *Step 2*: with another mismatched key $(\mathbf{A} + \varepsilon\mathbf{R}, \mathbf{I}_0)$, decrypt $\mathbf{x}(t)$ to get $\tilde{\mathbf{s}}(t)$ – an estimated version of $\mathbf{s}(t)$, where $\varepsilon \in (0, 1)$ and \mathbf{R} is a $P \times (P+Q)$ random $(1, -1)$ -matrix.

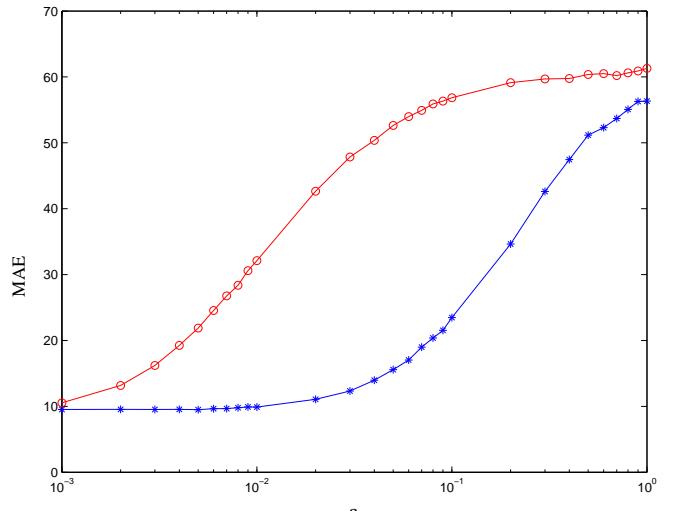
For each value of ε , the second step was repeated for 100 times to get a mean value of the recovery error (measured in MAE – mean absolute error)⁴. Then, we can observe the relationship between the recovery error and the value of ε . Figure 1 shows the experimental results when the plaintexts are a digital image and a speech file, respectively.

The experimental results confirms that a mismatched key can approximately recover the plaintext. Considering that humans have a good capability of resisting errors in images and speech, even relatively large errors may not be able to prevent a human attacker from recognizing the plain-image or plain-speech. Thus, the value of ε may be relatively large. When $P = 4$, $\mathbf{A} = [\mathbf{B}, \beta\mathbf{B}]$ and $\varepsilon = 0.1$, we give two examples of such recognizable plaintexts with relatively large errors in Figs. 2 and 3.

From the above experimental results, we can exhaustively search for an approximate version of \mathbf{A} under the finite precision $\varepsilon = 0.01 \sim 0.1$. Such an approximate version of \mathbf{A} is then used to roughly reveal the plaintext. Considering the searching complexity is $O(\varepsilon^{-(P+Q)})$, such an exhaustive search is feasible when P, Q is not very large⁵. When $P =$

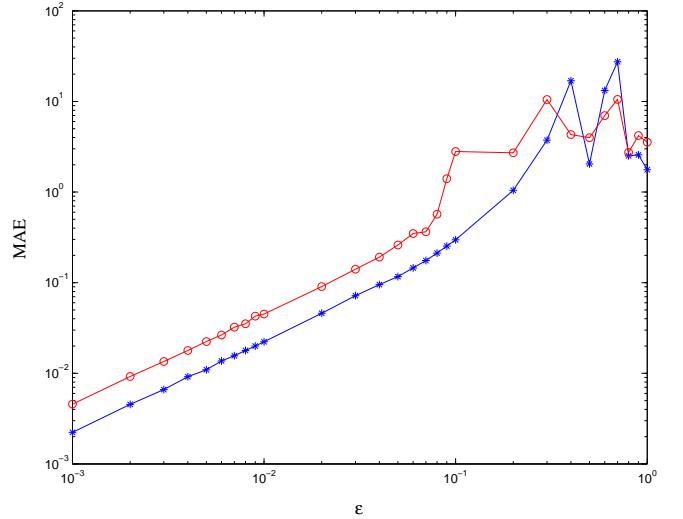
⁴When the plaintext is a digital image with 256 gray scales, we first calibrate each sub-image into the range $\{0, \dots, 255\}$ and then calculate the recovery error of the whole image.

⁵In [31]–[37], small values are used in all examples: $P = 2$ or 4 and $Q \leq P$.



Legend: * – $P = Q = 4$; o – $P = 4$ and $\mathbf{A} = [\mathbf{B}, \beta\mathbf{B}]$
($\beta = 10$).

a)



Legend: * – $P = Q = 4$; o – $P = 4$ and $\mathbf{A} = [\mathbf{B}, \beta\mathbf{B}]$
($\beta = 2$).

b)

Fig. 1. The experimental relationship between the recovery error and the value of ε : a) the plaintext is a digital image "Lenna" (Fig. 3a); b) the plaintext is a speech file "one.wav" that corresponds to the pronunciation of the English word "one" (from Merriam-Webster Online Dictionary, <http://www.m-w.com>).

2 and $\mathbf{A} = [\mathbf{B}, \beta\mathbf{B}]$, we carried out a large number of experiments in the following steps:

- *Step 1*: for a randomly-generated key $(\mathbf{B}, \mathbf{I}_0)$, calculate the ciphertext $\mathbf{x}(t)$ corresponding to a plaintext $\mathbf{s}(t)$;
- *Step 2*: randomly generate a matrix \mathbf{R} (each element over the interval $[-1, 1]$), and then decrypt $\mathbf{x}(t)$ with the guessed key $(\mathbf{R}, \mathbf{I}_0)$ to get $\tilde{\mathbf{s}}(t)$;
- *Step 3*: repeat *Step 2* for r rounds, output the recovered plaintext $\tilde{\mathbf{s}}^*(t)$, every segment of which corresponds to the best recovery performance in all the r rounds;
- *Step 4*: for the i -th segment of $\tilde{\mathbf{s}}^*(t)$, find the corresponding matrix \mathbf{R} , extract its i -th row of its inverse \mathbf{R}^{-1} to form the i -th row of $\tilde{\mathbf{B}}^{-1}$, the inverse of an estimation

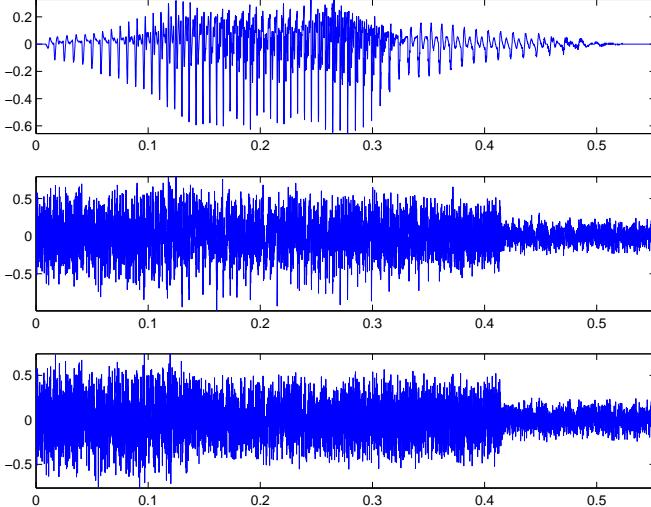


Fig. 2. An example of human capability against large noises in speech. From top to bottom: the original plain-speech “one.wav”, the recovered speech, the recovery error (MAE=0.164103). For reader’s sake, the recovered speech is posted online at http://www.hooklee.com/Papers/Data/BSSE/one_MAE=0.164103.wav.

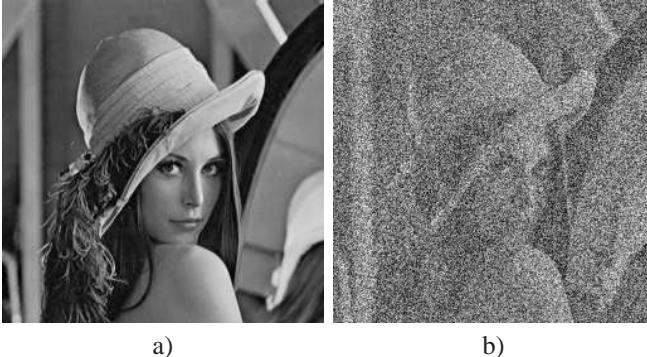


Fig. 3. An example of human capability against large noises in images: a) the original plain-image “Lenna”; b) the recovered image (MAE=47.6913).

of the original matrix \mathbf{B} .

Assuming that the target finite precision is $\varepsilon > 0$, the interval $[-1, 1]$ is divided into $n_\varepsilon = \lceil 2/\varepsilon \rceil$ sub-intervals. Without loss of generality, assuming that $2/\varepsilon$ is an integer, then each sub-interval is of equal size. Thus, if the element in the random matrix \mathbf{R} has a uniform distribution over $[-1, 1]$, the probability that $|r_{i,j} - a_{i,j}| < \varepsilon$ occurs at least one time in r rounds of experiment is $p(n_\varepsilon, r) = 1 - (1 - 1/n_\varepsilon)^r$, where $r_{i,j}$ and $a_{i,j}$ are the (i, j) -th elements of \mathbf{R} and \mathbf{A} , respectively. One can easily deduce that $p(n_\varepsilon, r)$ is an increasing function with respect to r and

$$\begin{aligned} p(n_\varepsilon, n_\varepsilon) > \lim_{n_\varepsilon \rightarrow \infty} p(n_\varepsilon, n_\varepsilon) &= 1 - \lim_{n_\varepsilon \rightarrow \infty} (1 - 1/n_\varepsilon)^{n_\varepsilon} \\ &= 1 - e^{-1} \approx 0.6321, \end{aligned}$$

which leads to the result that $p(n_\varepsilon, r) > 1 - e^{-1}$ when $r \geq n_\varepsilon$. In other words, with $r \geq n_\varepsilon$ experiments, it is a high-probability event that we have at least one $r_{i,j}$ “equal” to $a_{i,j}$ under the finite precision ε . To get an approximate estimation of the i -th row of \mathbf{A} , we can see that $r = O(n_\varepsilon^P)$ rounds of experiment are needed.

Apparently, the above steps actually simulate the process of a real ciphertext-only attack that tries to reveal the plaintext and to exhaustively guess \mathbf{B}^{-1} (under the assumption that \mathbf{I}_0 has been known). Note that MAE cannot be calculated to evaluate the recovery performance in a real attack, in which one does not know the plaintext. Fortunately, exploiting the large information redundancy existing in natural images and speech, one can turn to use some other measures to reflect the recovery performance of each segment of $\tilde{\mathbf{s}}(t)$. In our experiments, we use a measure called MANE (mean absolute neighboring error), which is defined as follows for the i -th segment of $\tilde{\mathbf{s}}(t)$

$$\frac{1}{T-2} \sum_{t=2}^{T-1} \frac{|\tilde{s}_i(t) - \tilde{s}_i(t-1)| + |\tilde{s}_i(t) - \tilde{s}_i(t+1)|}{2}, \quad (9)$$

where T denotes the segment length. In Figs. 4 and 5, one recovered plain-speech and two recovered plain-images are shown for demonstration. One can see that $r = O(10,000)$ (or $\varepsilon \approx 0.01$) is sufficient to get a good estimation of the plaintext.

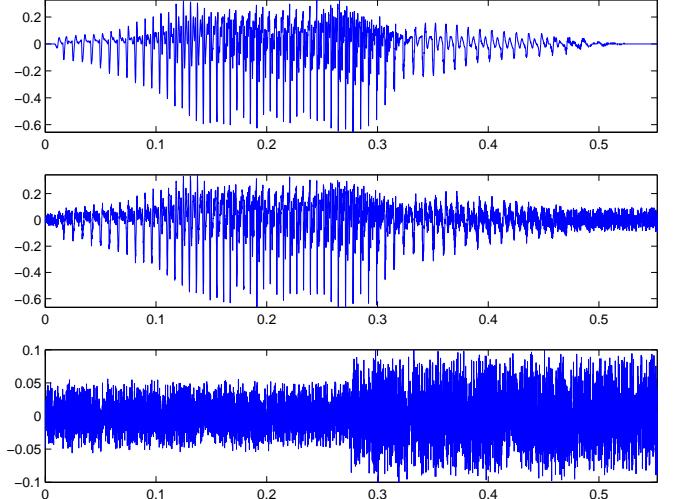
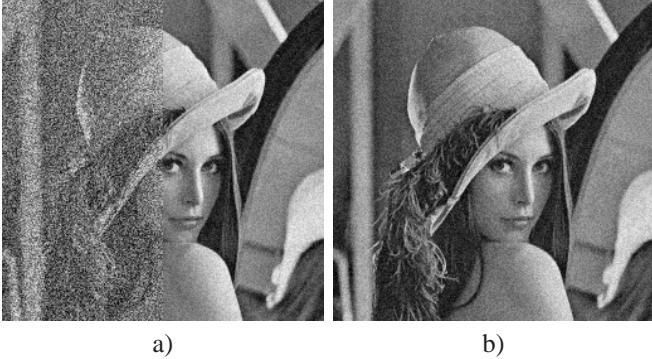


Fig. 4. A recovered speech in one 50,000-round experiment of exhaustively guessing \mathbf{A} when $P = 2$ and $\mathbf{A} = [\mathbf{B}, \beta\mathbf{B}]$. From top to bottom: the original plain-speech “one.wav”, the recovered speech (MANE of each segment: 0.0469, 0.0521), the recovery error. For reader’s sake, the recovered speech is posted online at http://www.hooklee.com/Papers/Data/BSSE/one_MANE=0.0469-0.0521.wav.

Note that for 2-D images the above 1-D MANE may be generalized to include more neighboring pixels, thus achieving a more accurate description of the recovery performance. In addition, multiple quality factors can be employed to further increase the efficiency of evaluation of the recovery performance.

3) *Low Sensitivity to $\mathbf{k}(t)$* : Due to the same reason of the low sensitivity to \mathbf{A} , one can deduce that the BSS-based encryption scheme is also insensitive to the key signal $\mathbf{k}(t)$. Given two key signals $\mathbf{k}_1(t)$ and $\mathbf{k}_2(t)$, if the maximal difference of all elements is ε , each element of $|\mathbf{A}_k \mathbf{k}_1(t) - \mathbf{A}_k \mathbf{k}_2(t)|$ is not greater than $Q \max(|\mathbf{A}_k|) \varepsilon = Q\varepsilon$. Since $\mathbf{k}(t)$ itself is not part of the secret key, but generated from \mathbf{I}_0 , this problem



a) b)

Fig. 5. Two recovered plain-images in our experiments of exhaustively guessing \mathbf{B} when $P = 2$ and $\mathbf{A} = [\mathbf{B}, \beta\mathbf{B}]$: a) $r = 1,000$ (MANE of each segment: 39.7491, 14.9373); b) $r = 10,000$ (MANE of each segment: 16.3888, 15.1722).

does not have much negative influence on the security of the whole cryptosystem against ciphertext-only attacks.

4) *Low Sensitivity to Plaintext*: Another cryptographical property required by a good cryptosystem is that the encryption is very sensitive to plaintext, i.e., the ciphertexts of two plaintexts with a slight difference should be much different [12]. However, this property does not hold for the BSS-based encryption scheme. Given two key signals $\mathbf{s}_1(t)$ and $\mathbf{s}_2(t)$, if the maximal difference of all elements is ε , each element of $|\mathbf{A}_s \mathbf{s}_1(t) - \mathbf{A}_s \mathbf{s}_2(t)|$ is not greater than $P \max(|\mathbf{A}_s|) \varepsilon = P\varepsilon$. When the same secret key is used to encrypt two close-correlated plaintexts, such as a plaintext and its watermarked version, this security defect means that the exposure of one plaintext leads to the revealment of both.

5) *Differential Attack*: Given two plaintexts $\mathbf{s}^{(1)}(t)$ and $\mathbf{s}^{(2)}(t)$, if they are encrypted with the same key $(\mathbf{A}, \mathbf{I}_0)$, we can get the following formula from Eq. (3):

$$\Delta_{\mathbf{x}}(t) = \mathbf{A}_s \Delta_{\mathbf{s}}(t), \quad (10)$$

where $\Delta_{\mathbf{x}}(t) = \mathbf{x}^{(1)}(t) - \mathbf{x}^{(2)}(t)$ and $\Delta_{\mathbf{s}}(t) = \mathbf{s}^{(1)}(t) - \mathbf{s}^{(2)}(t)$. Note that $\mathbf{A}_k \mathbf{k}(t)$ disappears in the above equation. This means that from the differential viewpoint only \mathbf{A}_s is the secret key, i.e., \mathbf{I}_0 is removed from the key. Considering the low sensitivity of the encryption scheme to \mathbf{A} , under finite precision ε the key space becomes $O(P\varepsilon^{-P})$, and one might exhaustively search \mathbf{A}_s to recover the plaintext differential as follows:

$$\Delta_{\mathbf{s}}(t) = \mathbf{A}_s^{-1} \Delta_{\mathbf{x}}(t). \quad (11)$$

From the obtained plaintext differential, one can get a mixed view of the two interested plaintexts, from which both plaintexts may be completely recognizable by humans. See Figs. 6 and 7 for four plaintext differentials of two speech files and two images.

Denoting the guessed matrix by $\tilde{\mathbf{A}}_s$, we have

$$\tilde{\Delta}_{\mathbf{s}}(t) = \tilde{\mathbf{A}}_s^{-1} \Delta_{\mathbf{x}}(t) = \tilde{\mathbf{A}}_s^{-1} \mathbf{A}_s \Delta_{\mathbf{s}}(t). \quad (12)$$

Apparently, if $\tilde{\mathbf{A}}_s \neq \mathbf{A}_s$, the obtained plaintext differential $\tilde{\Delta}_{\mathbf{s}}(t)$ will have an inter-segment mixture, which may make the recognition of the two plaintexts more difficult. Fortunately, when P is relatively small, such an inter-segment mixture may not be too severe to prevent the recognition of the

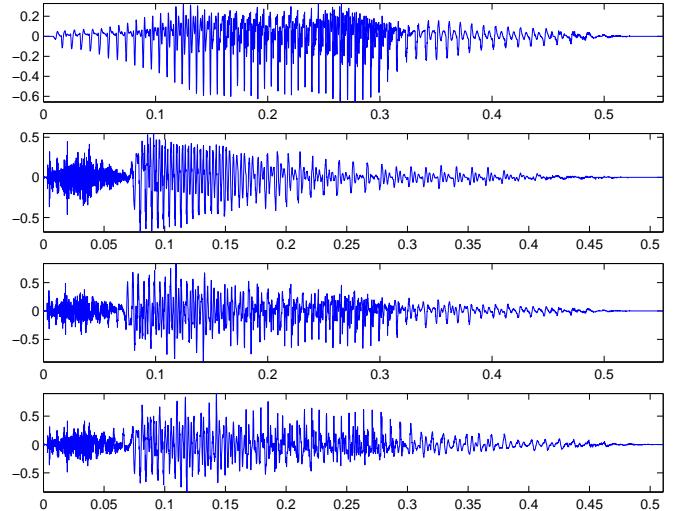


Fig. 6. Differentials of two plain-speech files. From top to bottom: the first speech “one.wav”, the second speech “two.wav”, the differential one-two, the differential two-one. For readers’ sake, the two differential speech files are posted online at <http://www.hooklee.com/Papers/Data/BSSE/one-two.wav> and <http://www.hooklee.com/Papers/Data/BSSE/two-one.wav>.



Fig. 7. Differentials of two plain-images, “Lenna” and “cameraman”: a) Lenna-cameraman; b) cameraman-Lenna.

two plaintexts by humans. More importantly, our experiments showed that humans can even be able to recognize the two plaintexts even when the mismatch between \mathbf{A}_s and $\tilde{\mathbf{A}}_s$ is not very small. When $P = 2$,

$$\mathbf{A}_s = \begin{bmatrix} 0.7123 & -0.4272 \\ 0.1958 & 0.1295 \end{bmatrix}, \quad \tilde{\mathbf{A}}_s = \begin{bmatrix} 0.5914 & 0.9527 \\ 0.5726 & 0.1437 \end{bmatrix}, \quad (13)$$

a plaintext differential obtained in our experiments is shown in Fig. 8. One can see that both plain-images, “Lenna” and “cameraman”, can still be roughly recognized from such a heavily mixed differential. Another obtained plain-speech differential for “one.wav” and “two.wav”, is shown in Fig. 9, from which the two English words (“one” and “two”) are also perceptible.

In this differential attack, the quality evaluation factors (such as MANE) used in Sec. III-A.2 is not suitable to automatically determine the best result in many plaintext differentials, because each segment of the obtained plaintext differential is also a natural signal with abundant information redundancy. Instead, one has to output all obtained differentials, and check them with naked eyes or ears to find a perceptually-optimal



Fig. 8. One obtained plain-image differential when \mathbf{A}_s and $\tilde{\mathbf{A}}_s$ have a relatively large mismatch as shown in Eq. (13).

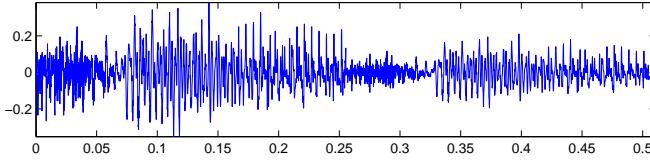


Fig. 9. One obtained plain-speech differential when \mathbf{A}_s and $\tilde{\mathbf{A}}_s$ have a relatively large mismatch. For readers' sake, this differential speech is posted online at <http://www.hooklee.com/Papers/Data/BSSE/two-one-large-mismatch.wav>.

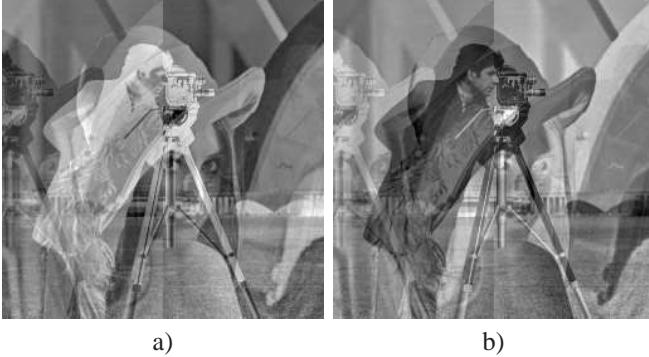


Fig. 10. A visually-optimal result obtained in 100 plain-image differentials: a) the differential; b) the negative image of the differential.

result with the least inter-segment mixture. Figure 10 shows such a result in 100 plain-image differentials when $P = 2$ and \mathbf{A} follows Eq. (13). By checking each segment separately and combine the P optimal segments together, one can further get a better result with less inter-segment mixture.

While this differential attack works well for $P = 2$ as shown above, it will become infeasible when P is sufficiently large, due to the following facts: 1) the inter-segment mixture is too severe; 2) the complexity of checking all $O(\varepsilon^{-P})$ differentials is beyond humans' capability.

B. Known-Plaintext Attack

In this kind of attack, one can access to a number of plaintexts that are encrypted with the same key. Then, from Eq. (10), with P plaintext differentials, one immediately knows that the mixing matrix can be uniquely determined as follows:

$$\mathbf{A}_s = \Delta_{\mathbf{X}}(t)(\Delta_{\mathbf{S}}(t))^{-1}, \quad (14)$$

where $\Delta_{\mathbf{S}}(t)$ and $\Delta_{\mathbf{X}}(t)$ are $P \times P$ matrices, constructed row by row from the P plaintext differentials and the corresponding ciphertext differentials, respectively. Then, $\mathbf{A}_k \mathbf{k}(t)$ can be further solved from any plaintext and its ciphertext:

$$\mathbf{A}_k \mathbf{k}(t) = \mathbf{x}(t) - \mathbf{A}_s \mathbf{s}(t). \quad (15)$$

Now, $(\mathbf{A}_s, \mathbf{A}_k \mathbf{k}(t))$ can be used to recover other plaintexts encrypted by the same key $(\mathbf{A}, \mathbf{I}_0)$. Note that $\mathbf{A}_k \mathbf{k}(t)$ has a finite length determined by the maximal length of all known plaintexts, so $(\mathbf{A}_s, \mathbf{A}_k \mathbf{k}(t))$ can only recover plaintexts under this finite length.

When $\mathbf{A} = [\mathbf{B}, \beta \mathbf{B}]$, the key signals can also be determined:

$$\mathbf{k}(t) = \frac{\mathbf{s}(t) - \mathbf{B}^{-1} \mathbf{x}(t)}{\beta}. \quad (16)$$

If the PRNG used is not cryptographically strong (such as LFSR [12]), it may be possible to further derive the secret seed \mathbf{I}_0 , thus completely breaking the BSS-based encryption scheme.

Note that n distinct plaintexts can generate $\binom{n}{2} = n(n-1)/2$ plaintext differentials. Solving the inequality $n(n-1)/2 \geq P$, one can get the number of required plaintexts to yield at least P plaintext differentials:

$$n \geq \lceil \sqrt{P - 1/4} + 1/2 \rceil \approx \sqrt{P}. \quad (17)$$

C. Chosen-Plaintext/Ciphertext Attack

In chosen-plaintext attack, one can freely choose a number of plaintexts and observe the corresponding ciphertexts, while in chosen-ciphertext attack, one can freely choose a number of ciphertexts and observe the corresponding plaintexts. So in these attacks, one can choose P plaintext differentials easily, which means that the above differential known-plaintext attack still works in the same way.

IV. DISCUSSION

As we pointed out in last section, the BSS-based encryption scheme is always insecure against plaintext attack. So the secret key cannot be repeatedly used in any case. This means that the encryption scheme has to work like a common stream cipher, by changing the secret key for each distinct plaintext. However, in this case, $\mathbf{k}(t)$ (equivalently, the secret seed \mathbf{I}_0) is enough to provide a high level of security, since $\mathbf{k}(t)$ satisfies the cryptographical properties in a perfectly secure one-time-a-pad cipher (see Sec. V.B of [37]). Then, the mixing matrix \mathbf{A} becomes excessive.

Even when one wants to add a second defense to potential attacks by applying the BSS mixing, the low sensitivity of encryption/decryption to the mixing matrix \mathbf{A} (recall Sec. III-A.2) makes this goal less useful. As a result, with the current encryption design, the BSS model does not play a key role in the security of the scheme. The real core of the encryption scheme is the embedded PRNG that is in charge of generating the key signals masking the plaintexts.

If one wants to use the BSS-based encryption scheme with repeatedly used key, some essential modifications have to be made to reinforce the security against various attacks.

Following the cryptanalytic results given in last section, we suggest adopting two countermeasures simultaneously: 1) use a sufficiently large P ; 2) like the design of most modern block ciphers [12], iterate the BSS-based encryption for many rounds to avoid the original scheme's low sensitivity to the secret key and plaintext. It is obvious that both countermeasures will significantly influence the encryption/decryption speed of the encryption scheme. It seems doubtful if such an enhanced encryption scheme will have any advantages compared with other multiple-round block ciphers, especially AES [13] that can be optimized to run with a very high rate on PCs [40].

Finally, it deserves mentioning that the original BSS-based encryption scheme can be used to realize **lossy** decryption, an interesting feature that may find useful in some real applications⁶. This feature means that an encryption scheme can still (maybe roughly) recover the plaintext even when there are some errors in the ciphertexts. A typical use of this feature is that the ciphertext can be compressed with some lossy algorithms to save the required storage in local computers or the channel width for transmission. For the BSS-based encryption scheme, the lossy decryption feature is ensured by low sensitivity of decryption to ciphertext, which is due to the same reason of the low sensitivity of encryption to plaintext (recall Sec. III-A.4). However, keep in mind that the lossy decryption feature is induced by the low sensitivity to plaintext/ciphertext, so there is a tradeoff between this feature and security.

V. CONCLUSION

This paper analyzes the security of an image/speech encryption scheme based on BSS mixing technology [31]–[37]. It has been shown that this BSS-based encryption scheme suffers from some security defects, including its vulnerability to a ciphertext-only differential attack, known/chosen-plaintext attack and chosen-ciphertext attack. It remains an open problem how to apply BSS technology to construct cryptographically strong ciphers.

REFERENCES

- [1] H. J. Beker and F. C. Piper, *Secure Speech Communications*. London: Academic, 1985.
- [2] I. J. Kumar, “Cryptology of speech signal,” in *Cryptology: System Identification and Key-Clustering*. Laguna Hills, California: Aegean Park Press, 1997, ch. 6.
- [3] R. K. Nichols and P. C. Lekkas, “Speech cryptology,” in *Wireless Security: Models, Threats, and Solutions*. New York: McGraw-Hill, 2002, ch. 6, pp. 253–327.
- [4] B. Furht, D. Socek, and A. M. Eskicioglu, “Fundamentals of multimedia encryption techniques,” in *Multimedia Security Handbook*, B. Furht and D. Kirovski, Eds. Boca Raton, Florida: CRC Press LLC, 2004, ch. 3, pp. 93–132.
- [5] S. Li, G. Chen, and X. Zheng, “Chaos-based encryption for digital images and videos,” in *Multimedia Security Handbook*, B. Furht and D. Kirovski, Eds. Boca Raton, Florida: CRC Press LLC, 2004, ch. 4, pp. 133–167, preprint is available at <http://www.hooklee.com/pub.html>.
- [6] A. Uhl and A. Pommer, *Image and Video Encryption: From Digital Rights Management to Secured Personal Communication*. Boston: Springer Science + Business Media Inc., 2005.
- [7] B. Furht, E. Muharemagic, and D. Socek, *Multimedia Encryption and Watermarking*. Springer, 2005.
- [8] W. Zeng, H. Yu, and C.-Y. Lin, Eds., *Multimedia Security Technologies for Digital Rights Management*. Academic Press, 2006.
- [9] B. Javidi, *Optical and Digital Techniques for Information Security*. New York: Springer Science + Business Media Inc., 2005.
- [10] M. G. Kuhn, “Analysis for the nagravision video scrambling method,” Online document, available at <http://www.cl.cam.ac.uk/~mkg25>, 1998.
- [11] S. Li, C. Li, G. Chen, D. Zhang, and N. G. Bourbakis, “A general cryptanalysis of permutation-only multimedia encryption algorithms,” IACR’s Cryptology ePrint Archive: Report 2004/374, available online at <http://eprint.iacr.org/2004/374>, 2004.
- [12] B. Schneier, *Applied Cryptography – Protocols, Algorithms, and Source Code in C*, 2nd ed. New York: John Wiley & Sons, Inc., 1996.
- [13] National Institute of Standards and Technology (US), “Specification for the advanced encryption standard (AES),” Federal Information Processing Standards Publication 197 (FIPS PUB 197), November 2001.
- [14] J. Wen, M. Severa, W. Zeng, M. H. Luttrell, and W. Jin, “A format-compliant configurable encryption framework for access control of video,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 6, pp. 545–557, 2002.
- [15] S. Li, G. Chen, A. Cheung, and B. Bhargava, “On the design of perceptual mpeg-video encryption algorithms,” arXiv e-print, cs.MM/0501014, available at <http://arxiv.org/abs/cs.MM/0501014>, 2005.
- [16] M. Bertilsson, E. F. Brickell, and I. Ingemarsson, “Cryptanalysis of video encryption based on space-filling curves,” in *Advances in Cryptology – EuroCrypt’88*, ser. Lecture Notes in Computer Science, vol. 434, 1989, pp. 403–411.
- [17] J.-K. Jan and Y.-M. Tseng, “On the security of image encryption method,” *Information Processing Letters*, vol. 60, no. 5, pp. 261–265, 1996.
- [18] L. Qiao, K. Nahrstedt, and M.-C. Tam, “Is MPEG encryption by using random list instead of ZigZag order secure?” in *Proc. IEEE Int. Symposium on Consumer Electronics (ISCE’97)*, 1997, pp. 226–229.
- [19] T. Uehara and R. Safavi-Naini, “Chosen DCT coefficients attack on MPEG encryption schemes,” in *Proc. IEEE Pacific-Rim Conference on Multimedia (IEEE-PCM’2000)*, 2000, pp. 316–319.
- [20] C.-C. Chang and T.-X. Yu, “Cryptanalysis of an encryption scheme for binary images,” *Pattern Recognition Letters*, vol. 23, no. 14, pp. 1847–1852, 2002.
- [21] A. M. Youssef and S. E. Tavares, “Comments on the security of fast encryption algorithm for multimedia (FEA-M),” *IEEE Trans. Consumer Electron.*, vol. 49, no. 1, pp. 168–170, 2003.
- [22] S. Li and K.-T. Lo, “Security problems with improper implementations of improved FEA-M,” accepted by *J. Systems and Software*, preprint available online at <http://dx.doi.org/10.1016/j.jss.2006.05.002> and <http://arxiv.org/abs/cs.CR/0509036>, 2006.
- [23] S. Li and X. Zheng, “Cryptanalysis of a chaotic image encryption method,” in *Proc. IEEE Int. Symposium on Circuits and Systems*, vol. II, 2002, pp. 708–711.
- [24] ———, “On the security of an image encryption method,” in *Proc. IEEE Int. Conference on Image Processing*, vol. 2, 2002, pp. 925–928.
- [25] C. Li, S. Li, D. Zhang, and G. Chen, “Cryptanalysis of a chaotic neural network based multimedia encryption scheme,” in *Advances in Multimedia Information Processing - PCM 2004: 5th Pacific Rim Conference on Multimedia, Tokyo, Japan, November 30 - December 3, 2004. Proceedings, Part III*, ser. Lecture Notes in Computer Science, vol. 3333. Springer-Verlag, 2004, pp. 418–425.
- [26] C. Li, S. Li, G. Chen, G. Chen, and L. Hu, “Cryptanalysis of a new signal security system for multimedia data transmission,” *EURASIP J. Applied Signal Processing*, vol. 2005, no. 8, pp. 1277–1288, 2005.
- [27] C. Li, X. Li, S. Li, and G. Chen, “Cryptanalysis of a multistage encryption system,” in *Proc. IEEE Int. Symposium on Circuits and Systems*, 2005, pp. 880–883.
- [28] C. Li, S. Li, D.-C. Lou, and D. Zhang, “On the security of the Yen-Guo’s domino signal encryption algorithm (DSEA),” *J. Systems and Software*, vol. 79, no. 2, pp. 253–258, 2006.
- [29] S. Li, C. Li, G. Chen, and X. Mou, “Cryptanalysis of the RCES/RSES image encryption scheme,” IACR’s Cryptology ePrint Archive: Report 2004/376, available online at <http://eprint.iacr.org/2004/376>, 2004.
- [30] S. Li, C. Li, K.-T. Lo, and G. Chen, “Cryptanalysis of an image scrambling scheme without bandwidth expansion,” Cryptology ePrint Archive: Report 2006/215, available online at <http://eprint.iacr.org/2006/215>, 2006.
- [31] Q.-H. Lin and F.-L. Yin, “Blind source separation applied to image cryptosystems with dual encryption,” *Electronics Letters*, vol. 38, no. 19, pp. 1092–1094, September 2002.

⁶Another scheme is a matrix-based image scrambling system proposed in [41], as pointed out in [30].

- [32] Q. Lin and F. Yin, "Image cryptosystems based on blind source separation," in *Proceedings of the 2003 International Conference on Neural Networks and Signal Processing (ICNNSP'2003)*, vol. 2. IEEE, 2003, pp. 1366–1369.
- [33] Q.-H. Lin, F.-L. Yin, and Y.-R. Zheng, "Secure image communication using blind source separation," in *Proceedings of the IEEE 6th Circuits and Systems Symposium on Emerging Technologies: Frontiers of Mobile and Wireless Communication (CASSET'2004)*, vol. 1. IEEE, 2004, pp. 261–264.
- [34] Q. Lin, F. Yin, and H. Liang, "Blind source separation-based encryption of images and speeches," in *Advances in Neural Networks - ISNN 2005: Second International Symposium on Neural Networks, Chongqing, China, May 30 - June 1, 2005, Proceedings, Part II*, ser. Lecture Notes in Computer Science, J. Wang, X. Liao, and Z. Yi, Eds., vol. 3497. Heidelberg: Springer Berlin, 2005, pp. 544–549.
- [35] Q.-H. Lin, F.-L. Yin, and H.-L. Liang, "A fast decryption algorithm for BSS-based image encryption," in *Advances in Neural Networks - ISNN 2006: Third International Symposium on Neural Networks, ISNN 2006, Chengdu, China, May 28 - June 1, 2006, Proceedings, Part III*, ser. Lecture Notes in Computer Science, J. Wang, Z. Yi, J. M. Zurada, B.-L. Lu, and H. Yin, Eds., vol. 3973. Heidelberg: Springer Berlin, 2006, pp. 318–325.
- [36] Q.-H. Lin, F.-L. Yin, T.-M. Mei, and H. Liang, "A speech encryption algorithm based on blind source separation," in *Proceedings of the 2004 International Conference on Communications, Circuits and Systems (ICCCAS'2004)*, vol. 2. IEEE, 2004, pp. 1013–1017.
- [37] ———, "A blind source separation based method for speech encryption," *IEEE Trans. Circuits Syst. I*, vol. 53, no. 6, pp. 1320–1328, June 2006.
- [38] J.-F. Cardoso, "Blind signal separation: Statistical principles," *Proc. IEEE*, vol. 86, no. 10, pp. 2009–2025, 1998.
- [39] IEEE Computer Society, "IEEE standard for binary floating-point arithmetic," ANSI/IEEE Std. 754-1985, 1985.
- [40] B. Gladman, "AES and combined encryption/authentication modes," online document, available at <http://fp.gladman.plus.com/AES/index.htm>, 2006.
- [41] D. V. D. Ville, W. Philips, R. V. de Walle, and I. Lemanhieu, "Image scrambling without bandwidth expansion," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 6, pp. 892–897, 2004.